

Taller de limpieza de datos en Python

La limpieza y pre procesamiento de datos es una de las actividades más importantes dentro de la analítica de datos, puesto que la calidad de nuestros análisis y modelos depende en gran medida por no decir que en su totalidad de la calidad de los datos.

Con el propósito de poner en práctica las temáticas vistas se presenta un conjunto de datos fraccionado en diferentes archivos .csv. Usted deberá ‘limpiar’ estos archivos y unificarlos en un único conjunto de datos que sea de utilidad para análisis posteriores. Tenga en cuenta las siguientes recomendaciones:

1. Verifique el tipo de dato y modifíquelo si lo considera necesario
2. Verifique la consistencia en los valores asignados a la variable *tag*
3. Reemplace cada valor perdidos de *Bike Id* con "No disponible"
4. Realice una imputación Interpolar los valores perdidos en la columna *userCategory*
5. Reemplace cada valor perdidos en *EndStation Id* con "No disponible"

La descripción de las variables se muestra a continuación:

1. **Rental Id:** Un usuario puede alquilar cualquier bicicleta en la ciudad durante el día por un costo de £2. Esta columna corresponde al Id de una transacción de alquiler, por lo que puede corresponder a más de una bicicleta (No es el Id de un viaje)
2. **Duration:** Duración del viaje en segundos
3. **Bike Id:** Identificación de la bicicleta alquilada
4. **End Date:** Momento final del viaje
5. **EndStation Id:** Id de la estación donde el viaje finalizó
6. **Start Date:** Momento de inicio del viaje
7. **StartStation Id:** Id de la estación donde el viaje inicio
8. **tag** Etiqueta asignada a cada uno de los viajes para facilitar la agrupación de viajes
9. **userCategory** Puede ser: A (Usuario ocasional) o B (Usuario frecuente)